人类嗓音加工的神经机制
——来自正常视力者和盲人的脑神经证据

明莉莉[1,2] 胡学平[1,2,3*]

([1]江苏师范大学语言科学与艺术学院，徐州 221009)

([2]江苏省语言与认知神经科学重点实验室，语言能力协同创新中心，徐州 221009)

([3]南京师范大学文学院，南京 210097)

摘 要 人类嗓音作为听觉环境的重要组成部分，包含了大量副语言信息以帮助识别个体身份，尤其对于盲人而言，视觉面孔经验的缺乏使得嗓音信息成为其感知对方个体特征的主要来源。通过结合正常视力者和盲人在嗓音选择性加工和嗓音身份加工方面的研究，尝试揭示普遍的人类嗓音加工模式以及盲人特异的嗓音加工机制。此外，与视觉面孔加工相关的梭状回脑区也参与并卷入了两类人群的嗓音加工任务，以此为基础梳理的视听整合模型和基于去掩蔽效应的跨模态重组假说，为这两类人群嗓音加工神经基础的差异性提供了相应的解释机制，未来研究希望可以进一步探究盲人的嗓音加工策略，以及考察左侧颞上沟/回脑区在嗓音加工中的作用。

关键词 嗓音身份加工，正常视力者，盲人群体，梭状回，面孔加工

1 引言

嗓音(voice)是人和动物以发声器官为基础连续动作产生的用以传达信息或通讯的声音(Candiotti et al., 2013)，至少三千万年以前已经存在于脊椎动物的听觉环境中(Belin et al., 2011)，这一能力在物种演变中得到长期保留，具有十分广泛的进化意义。在人类社会发展过程中，嗓音逐渐成为语音(speech)的载体，但是嗓音的意义并不局限于“简单”的语音(Belin et al., 2004)，因为我们不仅能从嗓音中获得语言信息，还能提取大量的生物、心理及社会特征等副语言信息以帮助标记、区别和识别个体身份(Belin et al., 2011; Laver, 1980)，例如听者可以基于说话人嗓音物理特征的某些线索，从而对其性别、年龄、健康程度、外貌(如身高、 体重和面部特征)和外形尺寸等进行判断，因此关于嗓音本质和嗓音加工的相

*收稿日期：2021-02-23

[*] 国家自然科学基金青年科学基金项目(批准号：31900750)、中国博士后科学基金资助项目(资助编号：2020M671530)、江苏师范大学自然科学研究基金项目博士学位教师科研支持项目(资助编号：18XLRX011)和江苏师范大学校级科研与实践创新项目(资助编号：2020XKT851)等资金资助。

通讯作者：胡学平，E-mail: huxpxp@163.com

关研究具有十分重要的社会和生态学意义(Schweinberger et al., 2014)。

相比正常视力者，盲人群体的嗓音感知能力则显得尤其重要。盲人由于失明，其使用视觉面孔识别说话人的能力存在长期发展性缺失，从而嗓音成为这一群体感知说话人身份的主要信息来源，并且大量基于学习(名字-嗓音)-再认范式的研究证明盲人具有优于常人的嗓音身份再认能力，先天(出生时或出生后一个月内失明)和晚期盲人(16 岁以后失明，与早期盲人的概念相对，目前基于早期盲人的相关研究主要选取先天盲人群体和少数 2 岁内失明的被试)熟悉嗓音身份的速度更快(Focker et al., 2012; Focker et al., 2015; Holig et al., 2014a, 2014b)，嗓音身份再认正确率更高(尤其是先天盲人) (Braun, 2012; Bull et al., 1983; Focker et al., 2012; Holig et al., 2014a; Pang et al., 2020)，再认需要的反应时间更短(Focker et al., 2012; Focker et al., 2015)。这些以盲人群体为对象的嗓音方面的研究，不仅有助于我们了解盲人相比于常人的嗓音加工优势以及二者在神经机制上的异同，同时也以一种独特视角重新阐述了人类嗓音加工这一认知活动，为探究普遍的人类嗓音身份加工机制提供了一些借鉴意义。

本文将结合正常视力者和盲人群体的嗓音加工研究，从嗓音选择性加工和嗓音身份加工两个角度阐述人类嗓音感知加工的相关机制，然后讨论梭状回脑区在嗓音加工中的作用以及盲人面孔加工脑区的跨模态重组现象，最后就盲人的嗓音加工策略、左侧颞上沟脑区在嗓音加工中的作用等方面进行展望。

2 人类嗓音选择性加工的脑机制

人类嗓音是我们听觉环境中最重要的声音，它不仅以语音形式体现于生活的各个方面，而且携带了较多副语言信息以促进人类社会交往和互动，目前大量研究已经发现相比于其他环境声音、动物发声，我们的某些大脑区域对人类嗓音产生更强的激活现象，即表现出对于人类嗓音的选择性加工(Fecteau et al., 2004)。

Belin 等人(2000)使用功能磁共振成像(functional magnetic resonance imaging, fMRI)技术首次发现了人类嗓音的选择性加工脑区——颞上沟(superior temporal sulcus, STS)，作者通过对比人类嗓音材料和与其能量相当的环境声音(例如自然声音、动物声音和机械声音)的大脑加工情况，发现无论是包含语音的嗓音(单词、非词、外语)，还是纯嗓音(例如笑、叹息等)材料，双侧颞上沟上岸(upper bank of STS)的神经元都比加工环境声音时具有更强的响应。Bestelmeyer 等人(2011)采用重复经颅磁刺激(repetitive transcranial magnetic stimulation, rTMS)技术进一步论证了 STS 与人类嗓音加工之间的因果关系，作者对于右侧 STS 和控制参考点——即右侧缘上回(supramarginal gyrus, SMG)分别进行刺激，结果发现相比于参考脑区，对右侧 STS 脑区的干预显著损坏了个体的人类嗓音感知能力(这里是指辨别人类嗓音和非人类嗓音)。此后大量关于嗓音加工的脑机制研究也不断证实了 STS 是人类嗓音选择加工的核心脑区这一结论(Mathias & von Kriegstein, 2014)，Belin 等学者将以 STS 为核心的嗓音

加工脑区命名为颞叶嗓音区(temporal voice areas, TVAs)，主要沿双侧 STG/S 分布(Belin & Grosbras, 2010)，值得一提的是，后续大量研究都显示嗓音感知具有右偏侧化倾向，即在右侧脑区的激活程度更高(Andics et al., 2010; Fecteau et al., 2004; Koeda et al., 2006; Sharda & Singh, 2012)。

盲人由于视觉剥夺主要依靠嗓音信息加工，这种相对的单一化加工通路使得盲人群体成为研究人类嗓音加工机制的一个非常好的视角。盲人群体和正常视力者都对于人类嗓音选择性地激活双侧 STS(Dormal et al., 2018)，但相比于正常视力者和晚期盲人，先天和早期盲人在左侧 STS/G 诱发出对于人类嗓音刺激的更强激活(Dormal et al., 2018; Gougoux et al., 2009)，在右侧 STS/G 则未报告存在显著差异(Gougoux et al., 2009)，因此在人类嗓音的选择性加工方面，盲人的偏侧化倾向可能弱于正常视力者群体，而这一倾向在嗓音身份、嗓音情绪、言语加工等嗓音信息加工的不同层面上均发现了类似的结论。首先，虽然双侧颞叶均存在显著激活，但右侧颞叶在说话人身份加工中起主导作用(Scott, 2019)，但盲人的嗓音身份研究不仅验证了右侧脑区的重要性，也凸显了左侧脑区对于盲人嗓音加工的重要作用盲人嗓音感知时诱发的左侧 STS 脑信号与陌生嗓音辨别成绩呈显著的正相关(Gougoux et al., 2009)，盲人左侧 STS 具有更高的嗓音身份分类准确率(Fairhall et al., 2017)。但这些结果并不意味着盲人嗓音加工呈左偏侧化倾向，因为其右侧 STS 神经活动与陌生嗓音辨别成绩也具有很强的相关性(Gougoux et al., 2009)，综合左侧和右侧 STS 的激活结果和脑活动-行为之间关联的研究结果，可以发现相比常人右偏侧化的嗓音身份加工倾向，盲人双侧 STS 均卷入至嗓音身份加工活动中。其次，在嗓音情绪加工方面，正常视力者仅在右侧 pSTS 对嗓音情绪产生响应，先天盲人的嗓音情绪表征在两个半球整体的响应强度均较高 (Fairhall et al., 2017)，并且早期盲人可以在检测嗓音情绪效价方面达到与常人类似的准确性，但并未表现出任一半球的偏侧化模式(Gamond et al., 2017)；在言语加工方面，先天盲人也存在更强的双侧加工效应，其在正常视力者语言加工中发现的典型左侧偏侧化现象明显减少(Lane et al., 2017)。因此无论是嗓音的选择性加工还是嗓音信息加工的不同层面上，盲人都表现出半球偏侧化效应的减少。最后，盲人群体半球偏侧化效应的减少可能与失明时间之间存在相关关系。Gougoux 等人(2009)的研究表明先天盲人左侧 STS 的激活程度高于晚期盲人，一定程度上说明嗓音选择性加工的偏侧化效应的减少程度与失明时间相关，而在言语加工层面，盲人大脑加工偏侧化效应的减少程度也与失明时间之间存在着正相关性(Atilgan et al., 2017; Lane et al., 2017; Pant et al., 2020)，例如成年后失明的盲人不同于先天盲人，仍表征出与常人类似的额颞语言网络的左侧化加工倾向(Pant et al., 2020)。

相比于正常视力者，先天和早期盲人的嗓音加工活动减少了特异半球的偏侧化倾向并表现出更具双侧化加工的特征，这一现象可能与盲人特异性的听觉脑区结构模式和听觉信息加工模式相关。在结构连接方面，先天盲人的双侧颞上平面(次级听觉皮层)中成对子脑区在皮层厚度和表面积上的相关性均明显高于正常对照组，说明先天失明严重影响了听觉

相关脑区的结构连接，并且发展为半球间形态更加协调的模式(Atilgan et al., 2017)；在听觉加工机制方面，患有先天无眼球症的盲人和早期盲人的内侧膝状体对于同侧和对侧听觉刺激的反应同样强烈(不同于正常视力者典型的对侧偏向)，表明其皮层-丘脑的连接发生了改变，即盲人对听觉输入表现出双边内侧膝状体以及双侧初级听觉皮层的响应(Coullon et al., 2015)；因此盲人听觉加工脑区结构和功能的特异性，可能在一定程度上解释了其嗓音加工的双侧化加工倾向。

除了典型的嗓音加工脑区，盲人视觉皮层也产生对于人类嗓音的激活现象，先天和晚期盲人对于嗓音刺激在枕叶诱发出更高的血氧反应(Holig et al., 2014a, 2014b; Topalidis et al., 2020)，但实际上盲人的枕叶脑区参与了听觉工作记忆、听觉空间定位、听觉运动、声音频率、听觉语言等广泛听觉任务的加工活动(Collignon et al., 2013; Collignon et al., 2011; Gudi-Mindermann et al., 2018; Lewald & Getzmann, 2013; Loiotile et al., 2019; Poirier et al., 2006; Rimmele et al., 2019; Voss et al., 2008; Voss et al., 2011; Watkins et al., 2013)。相比于正常视力者枕叶对于听觉刺激的血氧反应明显降低，先天盲人和晚期盲人的枕叶则显著提高了听觉诱发的血氧依赖反应(Collignon et al., 2013)，先天盲人还存在“初级听觉皮层(primary auditory cortex, PAC/A1)-初级视觉皮层(primary visual cortex, PVC/V1)”的皮质连接，因而增强了盲人枕叶中 V1 对听觉刺激的血氧反应(Klinge et al., 2010)，结合听觉输入明显增强盲人听觉和视觉皮层之间的 γ 波段同步性和相关性的证据(Pelland et al., 2017; Schepers et al., 2012)，盲人视觉皮层对于嗓音刺激的响应可能体现了盲人对于听觉输入的一般性加工。

总的来说，虽然基于嗓音加工障碍人群的研究表明嗓音感知能力可以独立于一般听觉能力而存在，但盲人嗓音加工的证据强调了二者之间的关系，即嗓音加工很可能受到听觉能力和听觉加工机制的影响，从而两类人群表现出不同的嗓音偏侧化加工倾向和视觉脑区的卷入与否的活动模式。

3 人类嗓音身份加工的脑机制

3.1 神经典型发育人群的嗓音身份加工研究

根据听者对于说话人嗓音的熟悉程度，可以将嗓音大致分为熟悉嗓音和不熟悉嗓音，熟悉嗓音研究一般涉及对其嗓音身份再认或识别，而不熟悉嗓音研究则一般涉及嗓音身份辨别 (Van Lancker & Kreiman, 1987; Van Lancker et al., 1989)。基于神经典型发育人群的脑成像研究已经证实不同熟悉程度的嗓音身份加工具有不同的认知神经基础，Belin 和 Zatore (2003)发现说话人身份适应性(聆听同一说话人产出的不同音节)会减弱 STS/G 前部(anterior STS, aSTS)的激活强度，而不会对言语适应(聆听不同说话人产出的相同音节)产生这一效应，说明 STS/G 前部与加工熟悉的说话人身份相关，并在后续的大量研究中进一步证明了 STS/G 的前部和中部(middle STS, mSTS)是熟悉嗓音身份再认的重要脑区(Aglieri et al., 2021; Belin & Zatorre, 2003; Bonte et al., 2014; Latinus et al., 2011; Schelinski et al., 2016; von

Kriegstein et al. 2003; von Kriegstein & Giraud, 2004)，而 STS/G 的后部(posterior STS, pSTS)则参与了不熟悉嗓音身份的辨别加工(Lattner et al., 2005; Rama & Courtney, 2005; von Kriegstein & Giraud, 2004; Warren et al., 2006)，例如 Warren 等人(2006)等人发现区分不同陌生嗓音会在 pSTS 诱发更强的激活。所以虽然右侧颞叶，尤其是 STS/G 相关脑区具有对于人类嗓音的选择性反应，但显示出对于不同熟悉程度嗓音身份加工的分离，即 STS/G 前部和中部区域更多地参与了熟悉嗓音身份加工，STS/G 后部参与了陌生嗓音身份加工。

嗓音感知网络不仅包括 TVA 这一核心区域，还包括由几个前额叶脑区(也被称为额叶嗓音区，frontal voice areas，FVAs)和皮质下结构(例如杏仁核)组成的“拓展”部分，这部分脑区由嗓音诱发的激活程度虽然较小但是非常明显(Pernet et al., 2015)。一些脑成像研究表明 FVAs 不仅仅对于嗓音信息特异性敏感，而且也参与了熟悉嗓音的身份加工(Aglieri et al., 2021; Aglieri et al., 2018; Holig et al., 2017; Zaske et al., 2017)。基于人类 STS 结构与功能连接的元分析结果表明 STS 前部与某些腹侧流区域表现出较强的共同激活，例如额下回眶部(inferior frontal gyrus pars orbitalis) (属于中部 FVAs) 和前颞叶(anterior temporal lobe, ATL)脑区 (Erickson et al., 2017)，其中 ATL 也被报告参与个人身份的多模态识别(von Kriegstein et al., 2006)，包括嗓音身份识别这一过程(Andics et al., 2010; Belin & Zatorre, 2003; Lattner et al., 2005; Nakamura et al., 2001; von Kriegstein et al., 2003)，因此 aSTS 与 ATL、IFG (inferior frontal gyrus)等腹侧流区域的功能连接存在于熟悉嗓音身份的加工中。

3.2 临床群体的嗓音身份加工研究

来自脑损伤病人的临床研究也显示熟悉和不熟悉嗓音是两种可分离的认知神经活动(Blank et al., 2014; Peretz et al., 1994; Stevenage, 2018; Van Lancker et al., 1988; Van Lancker & Kreiman, 1987)。Van Lancker 和 Kreiman(1987)、Peretz 等人(1994)发现了两类不同的嗓音失认症(phonagnosia)——陌生嗓音辨别障碍和熟悉(这里具体指著名嗓音)嗓音再认障碍，前者主要与任一半球的颞叶损伤相关，后者主要与右脑损伤相关，包括右下顶叶或右颞-顶叶皮层，接近于右侧 STS。先天嗓音失认症患者不存在脑区损伤状况，但也存在嗓音再认障碍，例如难以再认著名嗓音和学习再认新的嗓音，而分辨陌生嗓音的能力依然完好 (Garrido et al., 2009)，因此嗓音失认症患者不同的行为学症状以及不同的生理解剖结构支持了熟悉和不熟悉嗓音身份加工是两种不同的认知活动。此外，关于精神分裂患者和高功能孤独症谱系障碍(autism spectrum disorders, ASD)人群嗓音再认障碍的相关研究则进一步验证了基于神经典型发育人群嗓音再认的神经影像结果，例如存在幻听症状的精神分裂人群再认熟悉嗓音能力受损，并伴有右侧 STG 激活程度的显著降低 (Zhang et al., 2008)。患有嗓音再认障碍的高功能 ASD 人群 aSTS 脑区与熟悉嗓音再认行为表现之间并未表现出类似于常人的脑信号-行为表现间的相关性(Schelinski et al., 2016)，因而右侧 aSTS 的功能失调可能解释了这类人群的嗓音再认障碍。

除了嗓音身份加工障碍人群的研究，结合来自盲人的相关研究能够帮助我们更加深入了解人类嗓音身份加工机制，因为盲人依赖于嗓音加工说话人相关信息，脑成像结果也显示盲人更为依赖双侧 STS 脑区进行嗓音身份加工，Fairhall 等人(2017)使用了多体素分类分析方法(multi-voxel pattern analysis, MVPA)对比了先天盲人和正常视力者辨别陌生嗓音的大脑活动，结果显示盲人在左侧 STG 具有更高的嗓音身份分类准确率，说明盲人更为依赖 STS 加工嗓音身份信息。Gougoux 等人(2009)发现盲人双侧 pSTS 的激活强度与陌生嗓音辨别成绩之间具有较高的相关性，而在正常视力者中不存在这一现象，这一结果进一步了强调了关于 pSTS 参与不熟悉嗓音身份编码的结论。

此外，盲人嗓音加工的事件相关电位(event-related potential, ERP)研究显示了其在时间加工上的优势，相比于正常视力者和晚期盲人，先天盲人可以在早期感知加工阶段(100~160ms)就可以诱发更强的熟悉说话人匹配效应(说话人不一致条件的激活显著强于一致条件)(Focker et al., 2012; Focker et al., 2015)，说明先天盲人可以更早知觉说话人身份是否一致。而先天盲人在一般听觉刺激中已经表现出更早探测的优势，例如在听觉刺激诱发的中潜伏期响应脑电成分(例如 Pa、Nb、P1 等)和长潜伏期响应脑电成分(指的是 100ms 之后的脑电反应，例如 N1、P2)上的潜伏期都更短(Elbert et al., 2002; Manjunath et al., 1998; Naveen et al., 1997; Roder et al., 1996)，N1 振幅更大(Roder et al., 2007; Roder et al., 1996; Topalidis et al., 2020)和不应期更短(Roder et al., 1999)，说明先天盲人可以更早更有效地感知编码听觉刺激，这一能力可能解释了其在嗓音感知中所表现出的时间加工优势。

4 面孔加工相关脑区在嗓音加工中的卷入及其解释

4.1 正常视力者的“FFA-TVA”连接模式及其解释

一般认为视觉面孔和听觉嗓音分别作为加工说话人身份信息的两种不同通路，但正常视力者和盲人群体的嗓音加工研究显示了这两种模态之间又具有紧密的关系，具体是指两类人群的面孔加工相关脑区可以通过某些特定的方式参与嗓音加工任务。在正常视力者嗓音再认研究中，大量研究证明 TVA(temporal voice areas)和 FFA(fusiform face area)在结构和功能上均存在着连接通路(Blank et al., 2011; Blank et al., 2015; Blank et al., 2014; Schall et al., 2013; von Kriegstein et al., 2005)，由于 FFA 脑区又可以根据不同任务诱发的激活情况划分为以下两种功能脑区：在听觉环境下由熟悉嗓音再认任务诱发的 FFA 脑区即 cFFA(cross modal-FFA)，和在人类视觉面孔选择性加工任务中诱发的 FFA 脑区即 vFFA(visual-FFA) (Blank et al., 2011; von Kriegstein et al., 2008)，下文将分别阐述这两种功能脑区与嗓音加工脑区的连接模式。

一方面，有研究发现了关于嗓音再认过程中的 TVA 和 cFFA 的连接模式。面孔失认症患者经过实验前的(单模态)嗓音熟悉学习之后，虽然嗓音再认率明显差于常人，但其 FFA

脑区表征出对于熟悉嗓音信息的跨模态反应，并且 FFA(即 cFFA)与 STS 的功能连接强度与常人无异(von Kriegstein et al., 2006)，说明在没有任何视觉面孔信息提示的情况下(包括实验前的学习阶段和正式实验阶段)，cFFA 可以独立于面孔表征能力而响应于熟悉嗓音信息，并与 STS 构建功能连接作用于嗓音再认活动。

另一方面，还有一些研究也强调了熟悉嗓音再认过程中 vFFA 脑区的卷入以及 vFFA 和 TVA 的连接模式，这些研究使用(已知面孔)个体熟悉的嗓音(von Kriegstein et al., 2005)或者嗓音-面孔整体学习后的嗓音材料(von Kriegstein et al., 2008; Schall et al., 2013)，基于独立的视觉面孔定位任务中的 FFA 脑区(即 vFFA)作为感兴趣区，发现了其在嗓音再认过程中有所卷入(von Kriegstein et al., 2008; Schall et al., 2013)并与 STS 脑区存在功能连接模式(von Kriegstein et al., 2005)，说明在已知视觉面孔刺激的情况下，vFFA 脑区可能也参与了嗓音再认加工活动。此外，结合面孔失认症患者的证据也可以发现面孔信息对于嗓音再认具有重要作用，这类人群较弱的面孔表征能力使其难以通过面孔-嗓音的整体感知改善后期(单模态下)嗓音再认能力，也不存在行为与面孔加工脑区激活强度的相关性，而经过同样熟悉任务的正常对照组则可以提升嗓音再认行为成绩，说明面孔表征能力和对于面孔信息的加工可以有效促进个体身份加工或者嗓音再认加工(von Kriegstein et al., 2008)。同时，研究发现正常对照组行为成绩的提升与嗓音再认任务中 FFA(这里

vFFA，因为是基于独立的视觉面孔定位任务中的 FFA 脑区作为感兴趣区)的激活强度呈正相关，说明正常对照组在脑响应层面上也有效使用了面孔信息以促进嗓音再认，强调了 vFFA 在熟悉嗓音再认加工中的重要性(Maguinness & von Kriegstein et al., 2021； von Kriegstein et al., 2008)。因此视觉面孔加工脑区在嗓音再认过程中存在卷入现象并有效地促进了嗓音再认活动，下面我们将从以下三个方面详细阐述这一机制。

首先，视觉面孔信息可以在早期阶段调节嗓音身份的感知加工。Focker 等人(2011)首先让被试学习所有嗓音和面孔材料从而对实验中的说话人身份产生较高的熟悉程度，通过面孔或嗓音刺激(S1)对于第二个嗓音刺激(S2)的启动效应考察了身份再认的视听交互时间进程，结果发现单模态听觉(先后呈现两个嗓音刺激)的说话人不一致效应诱发了 270~530ms 的 ERP 负波，跨模态(先后呈现面孔、嗓音刺激)说话人不一致效应则可以在更早的时间窗内(100~140 ms)产生反应，说明视觉面孔传递的身份信息可以调节嗓音的早期感觉加工。其次，面孔心理表征可以加速嗓音再认活动，Schall 等人(2013)使用脑磁图(magnetoencephalography, MEG)技术发现相比于“职业(非面孔类图片)-嗓音”的控制条件，面孔-嗓音的前期学习经历使得在嗓音再认加工的 100ms 左右诱发了右侧 FFA 的激活，200ms 时听觉感觉成分 M200 的响应速度得到明显加快。虽然在此之前 von Kriegstein 等人(2005)已经证明了熟悉嗓音可以诱发面孔加工脑区的激活，来自运动引起的失明(motion-induced blindness, MIB)颅内脑电证据也表明，梭状回的神经元调节可能与视觉意识有关(Matsuzaki et al., 2012)，但 Schall 等人(2013)的研究结果则进一步显示面孔加工脑区在完成

听觉说话人识别(200ms 左右)之前就已经做出响应(Schall et al., 2015; Schweinberger, 2001)，并且大脑利用这种视觉表征加速了听觉中的嗓音早期感觉加工。最后，面孔信息不仅可以调节和加速嗓音再认的早期知觉过程，也有效促进了嗓音再认行为表现，即通过短时间的面孔-嗓音整体感知，正常视力者可以在行为表现和脑响应层面上有效使用面孔信息以提高嗓音再认能力(von Kriegstein et al., 2008)，这种来自面孔的促进作用在噪声条件下的嗓音再认研究中也得到进一步验证(Maguinness & von Kriegstein, 2021)。

总的来说，正常视力者的熟悉嗓音再认加工可以根据说话人面孔信息的已知与否划分为两种不同的“TVA-FFA”通路，在不涉及任何面孔信息的嗓音再认加工中，FFA 脑区可以对于熟悉嗓音产生响应即发生跨模态信息加工，并且构建其与嗓音加工脑区之间的功能连接；在已知说话人面孔的嗓音再认加工中，视觉面孔加工脑区至少在一定程度上卷入了这一过程，并可以利用熟悉面孔信息加速和促进了嗓音身份加工，符合视-听觉整合理论(Maguinness et al., 2018)，即嗓音加工和面孔加工系统可以在多个加工阶段产生交互作用，并且跨模态信息整合相比于单一模态的处理更加迅速和高效(Schweinberger et al., 2011)，这种交互作用和信息处理的高效性可能有助于优化个人身份识别加工。

4.2 盲人前部梭状回脑区在嗓音加工中的卷入及其解释

来自盲人群体嗓音加工的研究结果也强调了右侧梭状回在嗓音加工中的卷入，例如早期盲人的右侧前部梭状回脑区(anterior fusiform gyrus, aFG)诱发了对于人类嗓音的选择性反应(Gougoux et al., 2009)，并在这一加工过程中与左侧 STS 构建了功能连接(Dormal et al., 2018)，早期盲人和晚期盲人右侧 aFG 可以诱发对于熟悉嗓音身份的匹配效应 (Holig et al., 2014a, 2014b)，综合这些结果可以发现涉及盲人嗓音加工的 FG 脑区主要位于右侧 aFG。

aFG 不仅是视觉面孔的选择性加工脑区(Jonas et al., 2016)，同时也是熟悉面孔身份加工的一个关键节点(Jonas et al., 2015)。Jonas 等人(2016)利用颅内脑电(intracranial electroencephalographic, iEEG)技术记录了腹外侧枕-颞脑区对于面孔选择性反应加工诱发的局部场电位活动，结果发现不仅在右侧中部梭状回脑区(一般被认为是 vFFA 脑区的空间定位)，在右侧前部梭状回脑区等前颞叶脑区也出现了较高的与面孔选择性反应相关的神经信号。而关于面孔失认症的研究表明右侧 aFG 与面孔再认能力之间具有一定关系，相比于常人，面孔失认症患者的 aFG 的体积明显减少，并且这一体积的减少与著名面孔再认行为成绩的降低显著相关(Behrmann et al., 2007)，说明 aFG 脑区可能影响和参与了面孔再认活动。Jonas 等人(2015)的研究为这一推论提供了因果性证据，作者发现刺激癫痫患者(刺激之前患者在著名面孔再认测试、物品和人脸辨别测试的结果显示为正常)的右侧 aFG 可以诱发短暂性面孔失认症，即无法再认著名面孔，而视觉能力和人脸识别能力完好。

而在盲人群体中，参与常人面孔身份和面孔选择性加工的右侧 aFG 脑区可以响应于嗓音身份(Holig et al., 2014a, 2014b)和嗓音选择性加工(Gougoux et al., 2009; Dormal et al.,

2018)，因此盲人右侧 aFG 脑区可能发生了跨模态重组现象。来自聋人面孔加工的证据也进一步佐证了这种个体识别加工相关脑区上的跨模态重组现象，相比于常人，聋人右侧“颞叶嗓音加工脑区”(常人嗓音加工脑区的同源位置)诱发了更强的面孔选择性反应，这一神经活动与面孔加工行为成绩也具有显著正相关性(Benetti et al., 2017)，说明缺少听觉输入的聋人“嗓音加工脑区”有效地参与面孔加工，而盲人右侧 aFG 在嗓音加工任务中的卷入，可能也体现了这种神经可塑性变化。

针对盲人这一跨模态重组现象，则可以通过“去掩蔽(unmasking effect)”进行解释，掩蔽效应是指发生在传统的特定模态的皮层(例如视觉皮层、听觉皮层)内的跨模态抑制效应，以视觉皮层为例，人类和成年动物的视觉皮层存在来自皮质或者皮质下连接的听觉和触觉输入，但是这些输入会被主要的视觉输入所掩蔽，不能引起主观的非视觉感觉和枕叶激活(Kupers et al., 2006; Ptito et al., 2008； Qin & Yu, 2013)， fMRI 和正电子放射断层造影(positron emission tomography, PET)研究已经证明，一种感觉通道的刺激可以抑制其他感觉通道对相应脑区的反应 (Gougoux et al., 2005; Haxby et al., 1994; Johnson & Zatorre, 2005; Kawashima et al., 1995; Laurienti et al., 2002; Shulman et al., 1997)。由于缺乏视觉输入，盲人视觉皮层只能征用和增强已有的听觉或触觉输入以处理(非视觉)信息，这种现象被称为“去掩蔽效应”，例如盲人枕叶中的 V1 增强了与 A1 的结构和功能连接以执行听觉任务(Klinge et al., 2010)。

虽然目前掩蔽效应主要发生于传统特定模态的皮层中，但一些与个体识别加工相关的特异性脑区也具有类似表现，例如正常视力者的右侧嗓音加工区对于视觉面孔刺激表现为负激活状态(Benetti et al., 2017)，Benetti 等人(2018)使用概率纤维追踪技术检验了正常视力者右侧 TVA 与 V2/3 的结构连接，发现二者具有非常高的连接概率(≥80%)(Benetti et al., 2018)，也就是说右侧嗓音加工区接受来自 V2/3 的视觉输入，但掩蔽了视觉信息的加工，不会产生对于视觉面孔的响应。由此可见，这种跨模态抑制效应不仅仅发生在传统的视觉听觉皮层中，也发生在与个体识别加工的相关脑区中。

结合早期聋人面孔加工的研究将有助于解释这种基于去掩蔽效应的跨模态重组现象，早期聋人的嗓音加工脑区(这里具体

右侧 mSTS/G)响应于面孔选择性和面孔身份加工，关于宏观结构连接性的进一步研究结果显示，早期聋人“TVA-FFA”的连接比正常听人更弱(57%：86%)(Benetti et al., 2018)，说明感觉缺失人群并未表现出更强的“TVA-FFA”连接通路，而且相比于“TVA-FFA”，早期聋人的“V2/3-TVA”通路具有更高的连接强度，支持了聋人右侧 TVA 是基于已有视觉输入的基础发生了跨模态重组，从而执行面孔加工任务，这一结果进一步强调了关于长期发展性感觉缺失脑区的去掩蔽效应这一理论，并且这种去掩蔽效应也发生在与个体识别加工的相关脑区上。早期聋人“嗓音加工脑区”基于去掩蔽效应的跨模态重组这一证据，一定程度上支持了先天盲人“面孔加工脑区”可能也具有类似的解释机制，即通过对于听觉信

息的掩蔽消失和输入强化的方式加工嗓音信息。

那么这一跨模态重组脑区与其他嗓音加工脑区之间是如何整合信息和共同作用的?目前仅在人类嗓音选择性加工研究中发现盲人右侧 aFG 与左侧 STS 脑区存在功能连接(Dormal et al., 2018)，而在例如盲人嗓音身份加工等研究中还缺乏关于多个脑区之间感知整合方式的讨论。未来可以考虑结合更加多元的分析方法和信噪比更高的颅内脑电技术、经颅磁刺激等技术手段对于盲人“视觉”相关脑区参与嗓音任务的跨模态加工、以及跨模态脑区与典型嗓音加工脑区之间作用的方向性等问题开展更进一步的研究，以完善盲人嗓音加工脑机制研究的解释。

5 总结与未来研究展望

目前关于嗓音加工相关研究主要结合神经典型发育人群和伴有嗓音感知障碍的特殊人群(例如高功能 ASD、精神分裂患者)的证据，本文在此基础上加入嗓音感知优势群体(盲人)的研究进一步验证了在正常视力者中发现的右侧 STS 是对于人类嗓音选择性加工的重要脑区这一结论，并且通过梳理后发现盲人群体嗓音加工可能受到特异性听觉加工机制的影响表征出不同于常人的双侧化加工倾向和“视觉”皮层的响应及卷入。并且这些来自不同群体的研究也强调了不同熟悉程度的嗓音身份加工活动具有可分离的神经基础这一观点，其中值得一提的是，正常视力者的熟悉嗓音身份再认加工活动存在两种 “TVA-FFA”连接通路，包括 FFA 脑区的跨模态信息加工和视觉面孔信息的促进作用；而盲人梭状回前部脑区可能发生了跨模态重组以参与多项不同的嗓音加工活动，来自聋人 TVA 加工面孔刺激的证据进一步解释了这种基于去掩蔽效应的跨模态重组现象，强调了关于嗓音和面孔这两种模态加工之间的紧密关系，但仍存在很多问题值得未来的进一步研究。

第一，目前的研究虽然支持盲人具有更好的嗓音感知能力和不同于常人的嗓音加工脑机制，但并未回答关于盲人优势的身份感知能力(例如熟悉嗓音身份的速度更快)是否是由于对于某些声学参数具有更高敏感性这一问题，因为人类嗓音信息与其他声音相比包含了更为复杂的声学参数，其中时间频谱特征、基频等信息的改变都可以影响嗓音感知能力。已有研究似乎发现了盲人在一般听觉加工中存在优势的原因，例如盲人对于听觉频率的感知与常人不同(Huber et al., 2019; Watkins et al., 2013)，早期盲人对于不同频率的纯音感知具有更窄的神经“谐调(tunning)”(Huber et al., 2019)，先天盲人对于高中低不同频率的纯音感知激活了枕叶中 V5 /MT 的区域(一般被认为是听觉/视觉运动加工区)，但是在正常对照组中则未激活，说明盲人的这一脑区可以对于听觉信息进行早期加工(Watkins et al., 2013)。但是这些在一般听觉研究中的发现能否被概括为盲人嗓音身份加工优势的原因，尤其在熟悉嗓音再认方面，还有待进一步思考，因此未来研究可以就盲人对于人类嗓音不同声学参数的感知这一问题进行探究，以深入了解盲人的嗓音加工机制。

第二，由于声音的表征存在自上而下和自下而上两种加工方式，例如对于某些声学特

征进行更为细致的分析属于自下而上的加工策略，关于超快语音加工相关研究表明盲人相比于正常视力者可能更多采取自下而上的加工策略。Hertrich 等人(2018)的研究发现 IFG 脑区对于语音包络中音节起始模板(syllable onset template)具有显著的锁相反应。相比于盲人群体，正常视力者 IFG 脑区增加了由于语音训练引起的锁相活动，说明正常视力者增强了自上而下的加工，而盲人对于快速语音的加工可能更倾向于采取更加精细化的自下而上策略(Hertrich et al., 2018)，那么盲人对于嗓音的加工方式是否也采用此类策略？另一方面，听觉注意、认知控制等因素对嗓音加工会产生自上而下的影响，例如 Hu 等人使用母语和陌生语言的时间反转语音材料考察了嗓音辨别的语音熟悉度效应，发现随着语音不熟悉性的增加，嗓音识别网络(voice identity network, VIN)增强了与认知控制相关的额顶网络之间的功能连接，说明嗓音辨别活动会受到来自额顶网络自上而下的调控(Hu et al., 2017)，那么盲人是否是基于与认知控制脑区或者听觉注意相关网络的功能连接的方式增强了对于嗓音刺激的自上而下的加工？对于盲人基于哪种调控方式的增强从而具有异于常人的嗓音身份加工优势这一问题，可以成为未来后续研究中值得验证的一个方向。

第三，盲人相比于正常视力者在左侧 STS/G 具有更强的嗓音选择性反应，以及左侧 STS/G 脑区激活程度与嗓音辨别能力之间具有一定的相关性，这种相对的左侧脑区增强现象一方面可能与盲人特异性听觉机制相关，另一方面也提示了我们需要关注正常视力者左侧脑区在嗓音加工的作用。一些学者使用基于 fMRI 信号的 MVPA 分析方法发现了正常视力人群的左侧 STS/G 与嗓音加工活动有关，Bonte 等人(2014)的研究表明正常视力者的左侧颞上皮层的说话人分类准确率与辨别行为表现之间存在显著正相关性，Aglieri 等人(2021)等人通过让被试手指按键的方式识别出经过熟悉学习的嗓音，结果发现左侧 aSTS/G、同侧 IFG 脑区的分类准确率与个体再认成绩具有显著正相关性(Aglieri et al., 2021)。因此除了典型的右侧 TVA 之外，一些重要的相关拓展脑区也被逐渐证实参与了正常视力者的嗓音加工，例如左侧 STS/G 和 IFG 脑区，但这些脑区与右侧 TVA 之间存在怎样的关系以共同参与嗓音信息加工可能还需要开展更多的相关研究。

第四，在

涉及说话人面孔信息的嗓音身份加工任务中，两类人群表现出梭状回脑区不同位置的卷入——盲人梭状回前部区域和正常视力者梭状回中后部区域，这些脑区都与视觉面孔加工密切相关，可以诱发视觉面孔加工反应。但在没有提供说话人面孔信息的前提下，这种视觉面孔加工相关脑区的卷入，一方面可能反映了梭状回脑区对于嗓音信息进行的跨模态加工，但是为什么盲人的跨模态加工脑区不同于常人？这一问题目前并没有得到很好的回答，但就这两个脑区的功能来看，二者仍存在一些加工差异，例如刺激 aFG 脑区则无法再认著名面孔(包括无法回忆名字和背景语义知识)，但不会发生面孔扭曲现象(Jonas et al., 2015)，而刺激 FFA 脑区可以造成视觉面孔扭曲效应，但不影响著名面孔再认和视觉能力(Parvizi et al., 2012)，因此相比于 FFA 脑区，aFG 脑区的功能可能更倾向于面孔相关的语义

联系或语义记忆(Mion et al., 2010)，例如说话人的名字和相关背景语义知识，这些语义联系与说话人身份加工密切相关。因此对于缺乏视觉经验的盲人而言，即使知觉材料是熟悉嗓音，也只能更加依靠与语义记忆相关的 aFG 脑区的加工，所以相比于正常视力者，盲人群体加工嗓音加工过程更加需要 aFG 脑区的参与。但是这些推论目前还未得到相关实证研究的支持，并且这一解释还存在一定局限性，例如无法解释盲人嗓音选择性加工过程中 aFG 脑区的卷入，因此还需要针对盲人群体开展更多嗓音加工研究。另一方面视觉面孔加工脑区的卷入现象还可能说明这些梭状回脑区本质上属于多模态加工脑区，可以响应于其他非视觉信息，但以上两点都要求我们需要对于梭状回脑区结构和功能需要具有更加深入的挖掘和认识。

第五，Pascual-Leone 和 Hamilton (2001)基于掩蔽效应提出的关于大脑结构的元模态假说(metamodal hypothesis)为跨模态重组现象提供了理论解释，这一假说认为大脑实际上可能表征为一种用于执行特定功能或计算的元模态结构，与感觉输入方式无关，虽然脑区的功能独立于输入方式，但其使用何种模态进行表达则取决于不同输入之间竞争的结果，获得优胜的感觉输入对于这一功能的选择性反应不断增强，最终产生并行、独立地处理不同感觉信号的大脑结构。以视觉皮层为例，在正常视力者中来自视网膜的视觉信息为实现纹状体皮层的特定功能提供了最佳的信息来源，视觉输入从而成为其主要输入方式。对于失去视觉的盲人群体而言，这些区域的神经解剖结构增强了存在的听觉或触觉输入(去掩蔽)以执行高度相似的计算任务，表现为功能稳定(functional constancy)的跨模态重组现象，即利用不同的信息输入方式执行类似的功能。因此结合这一假说，盲人的面孔加工脑区可以基于对潜在听觉输入的强化以处理说话人相关的信息，但关于 FFA 与听觉脑区之间是否存在结构连接还有待未来研究的检验。

第六，一些脑成像研究也发现了一些高度功能特异性脑区独立于模态加工的证据，例如盲人可以通过触觉感知 3D 面孔在外侧梭状回脑区产生对于面孔的选择性激活(Ratan Murty et al., 2020)；视觉运动选择性加工的 V5 脑区(Wilms et al., 2005)发现了其在盲人群体中参与了听觉运动加工(Bedny et al., 2010)；与听觉定位加工相关的后部听觉野(posterior auditory field，PAF)参与了聋人视觉定位的加工(Lomber et al., 2010)；视觉词形区域(visual word form area, VWFA)与盲文阅读有关(Burton et al., 2002; Reich et al., 2011)。这些研究说明脑区可能具有独立于输入方式的功能特异性，例如 V5 的实际功能主要是加工物体的空间运动。不同于 Pascual-Leone 和 Hamilton(2001)关于结构决定功能的观点，Lomber 等人(2010)虽然认同感觉缺失脑区可以在维持原有功能的基础上改变模态的输入方式，其更强调这些功能本质上具有超模态性质，对于特定模态的特征(例如颜色，方向等)加工则不会发生这种跨模态重组现象(Lomber et al., 2010)，因此关于大脑结构、功能以及输入方式三者之间的关系还需要更多实验研究的探索，厘清这些关系也将有助于进一步解释面孔加工和嗓音加工的相互作用机制。

第七，嗓音身份作为听觉客体知觉特征的一部分，不仅具有特异性的神经加工机制，而且还需要进一步考虑嗓音身份与其他知觉特征(例如言语内容、情绪色彩、空间位置等信息)如何交互和整合，以及如何影响听觉客体的完好性等问题都显得更为重要，结合鸡尾酒会场景下的目标言语感知研究则将有助于进一步回答这一问题，研究发现当启动刺激和目标刺激的嗓音身份相同时，嗓音身份的一致性可以促进被掩蔽的目标言语的识别(Huang et al., 2010; Xu et al., 2008; Yang et al., 2007)。此外，在没有嗓音身份启动的情况下，听者对于目标言语说话人身份的熟悉性也可以减少言语信息掩蔽的干扰(Holmes et al., 2018; Holmes & Johnsrude，2021)，fMRI 研究表明这种嗓音身份熟悉性所引发的言语识别优势与目标言语的大脑皮层（颞中回和左侧颞上回后侧）信噪比的显著增加相关(Holmes & Johnsrude，2021)。因此，无论是嗓音身份的一致性还是熟悉性，均可以促进目标言语的感知编码。关于知觉客体的各个特征之间的捆绑机制将有助于进一步解释这一现象，即当嗓音身份知觉特征与已知(一致的)或内在(熟悉的)的嗓音身份知觉特征匹配时，可以强化嗓音身份特征的知觉加工，也就强化了其与对于目标言语特征之间的知觉捆绑，从而提高了目标言语客体的显现度(saliency)及其与掩蔽客体的知觉分离性，最后抑制掩蔽客体的表达并促进目标言语的识别(李量 等, 2017)。总的来说，未来研究需要在嗓音身份加工研究的基础上，考察其与其它嗓音信息的交互过程，对于这些问题的探讨将有助于更好地揭示听觉语流中多种线索的动态知觉过程，对于听觉加工、言语认知以及人工智能等领域都具有重要意义。

最后，盲人群体受到语言文字习得年龄、受教育程度、生活环境等诸多因素的影响，整体的认知水平与一般人群存在一定差异，也具有更大的个体差异性，因此提高被试间基本信息的匹配性可以帮助我们得到更为可靠的数据结果。鉴于目前国内缺乏具有广泛使用的盲人智力测试量表，未来研究可以考虑开发一套适用于国内盲人的可快速检测的智力测试量表。

参考文献

李量, 郑英君, 吴超, 黎绢花, 张畅芯, 陆灵犀. (2017). 在鸡尾酒会场景下利用去掩蔽知觉线索提高言语识别的脑网络机制. *心理科学进展, (12)*, 2099−2110.

Aglieri, V., Cagna, B., Velly, L., Takerkart, S., & Belin, P. (2021). FMRI-based identity classification accuracy in left temporal and frontal regions predicts speaker recognition performance. *Scientific Reports, 11*(1), Article 489. https://doi.org/doi: 10.1038/s41598-020-79922-7.

Aglieri, V., Chaminade, T., Takerkart, S., & Belin, P. (2018). Functional connectivity within the voice perception network and its behavioural relevance. *Neuroimage, 183*, 356−365.

Andics, A., McQueen, J. M., Petersson, K. M., Gal, V., Rudas, G., & Vidnyanszky, Z. (2010). Neural mechanisms for voice recognition. *Neuroimage, 52*(4), 1528−1540.

Atilgan, H., Collignon, O., & Hasson, U. (2017). Structural neuroplasticity of the superior temporal plane in early

and late blindness. *Brain and Language, 170*, 71–81.

Bedny, M., Konkle, T., Pelphrey, K., Saxe, R., & Pascual-Leone, A. (2010). Sensitive period for a multimodal response in human visual motion area MT/MST. *Current Biology, 20*(21), 1900–1906.

Behrmann, M., Avidan, G., Gao, F., & Black, S. (2007). Structural imaging reveals anatomical alterations in inferotemporal cortex in congenital prosopagnosia. *Cerebral Cortex, 17,* 2354–23563.

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*(4), 711–725.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129–135.

Belin, P., & Grosbras, M. H. (2010). Before speech: cerebral voice processing in infants. *Neuron, 65*(6), 733–735.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14*(16), 2105–2109.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309–312.

Benetti, S., Novello, L., Maffei, C., Rabini, G., Jovicich, J., & Collignon, O. (2018). White matter connectivity between occipital and temporal regions involved in face and voice processing in hearing and early deaf individuals. *Neuroimage, 179*, 263–274.

Benetti, S., van Ackeren, M. J., Rabini, G., Zonca, J., Foa, V., Baruffaldi, F., . . . Collignon, O. (2017). Functional selectivity for face processing in the temporal voice area of early deaf individuals. *Proceedings of the National Academy of Sciences, 114*(31), E6437–E6446.

Bestelmeyer, P. E., Belin, P., & Grosbras, M. H. (2011). Right temporal TMS impairs voice detection. *Current Biology, 21*(20), R838–R839.

Blank, H., Anwander, A., & von Kriegstein, K. (2011). Direct Structural Connections between Voice- and Face-Recognition Areas. *Journal of Neuroscience, 31*(36), 12906–12915.

Blank, H., Kiebel, S. J., & von Kriegstein, K. (2015). How the human brain exchanges information across sensory modalities to recognize other people. *Human Brain Mapping, 36*(1), 324–339.

Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: merging evidence from patients and healthy individuals. *Neuroscience & Biobehavioral Review, 47*, 717–734.

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *Journal of Neuroscience, 34*(13), 4548–4557.

Braun, A. (2012). Speaker Recognition Ability of Blind and Sighted Subjects. *The International Journal of Speech，Language and the Law, 19*(2), 159–187.

Bull, R., Rathborn, H., & Clifford, B. R. (1983). The voice-recognition accuracy of blind listeners. *Perception, 12*(2), 223–226.

Burton, H., Snyder, A. Z., Conturo, T. E., Akbudak, E., Ollinger, J. M., & Raichle, M. E. (2002). Adaptive changes in early and late blind: a fMRI study of Braille reading. *Journal of Neurophysiology, 87*(1), 589–607.

Candiotti, A., Zuberbuhler, K., & Lemasson, A. (2013). Voice discrimination in four primates. *Behavioural Processes, 99*, 67–72.

Collignon, O., Dormal, G., Albouy, G., Vandewalle, G., Voss, P., Phillips, C., & Lepore, F. (2013). Impact of blindness onset on the functional organization and the connectivity of the occipital cortex. *Brain, 136*(9), 2769–2783.

Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., & Lepore, F. (2011). Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. *Proceedings of the National Academy of Sciences, 108*(11), 4435–4440.

Coullon, G. S., Jiang, F., Fine, I., Watkins, K. E., & Bridge, H. (2015). Subcortical functional reorganization due to early blindness. *Journal of Neurophysiology, 113*(7), 2889–2899.

Dormal, G., Pelland, M., Rezk, M., Yakobov, E., Lepore, F., & Collignon, O. (2018). Functional Preference for Object Sounds and Voices in the Brain of Early Blind and Sighted Individuals. *Journal of Cognitive Neuroscience, 30*(1), 86–106.

Elbert, T., Sterr, A., Rockstroh, B., Pantev, C., Muller, M. M., & Taub, E. (2002). Expansion of the tonotopic area in the auditory cortex of the blind. *Journal of Neuroscience, 22*(22), 9941–9944.

Erickson, L. C., Rauschecker, J. P., & Turkeltaub, P. E. (2017). Meta-analytic connectivity modeling of the human superior temporal sulcus. *Brain Structure & Function, 222*(1), 267–285.

Fairhall, S. L., Porter, K. B., Bellucci, C., Mazzetti, M., Cipolli, C., & Gobbini, M. I. (2017). Plastic reorganization of neural systems for perception of others in the congenitally blind. *Neuroimage, 158*, 126–135.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage, 23*(3), 840–848.

Focker, J., Best, A., Holig, C., & Roder, B. (2012). The superiority in voice processing of the blind arises from neural plasticity at sensory processing stages. *Neuropsychologia, 50*(8), 2056–2067.

Focker, J., Holig, C., Best, A., & Roder, B. (2015). Neural plasticity of voice processing: Evidence from event-related potentials in late-onset blind and sighted individuals. *Restorative Neurology and Neuroscience, 33*(1), 15–30.

Focker, J., Holig, C., Best, A., & Roder, B. (2011). Crossmodal interaction of facial and vocal person identity information: An event-related potential study. *Brain Research, 1385*, 229–245.

Gamond, L., Vecchi, T., Ferrari, C., Merabet, L. B., & Cattaneo, Z. (2017). Emotion processing in early blind and sighted individuals. *Neuropsychology, 31*(5), 516–524.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., . . . Duchaine, B. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia, 47*(1), 123–131.

Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., & Zatorre, R. J. (2009). Voice perception in blind persons: a functional magnetic resonance imaging study. *Neuropsychologia, 47*(13), 2967–2974.

Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P., & Lepore, F. (2005). A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals. *PLoS Biology, 3*(2), 0324-0333.

Gudi-Mindermann, H., Rimmele, J. M., Nolte, G., Bruns, P., Engel, A. K., & Roder, B. (2018). Working memory training in congenitally blind individuals results in an integration of occipital cortex in functional networks. *Behavioural Brain Research, 348*, 31–41.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional

organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience, 14*(11), 6336–6353.

Hertrich, I., Dietrich, S., & Ackermann, H. (2018). Cortical phase locking to accelerated speech in blind and sighted listeners prior to and after training. *Brain and Language, 185*, 19–29.

Holig, C., Focker, J., Best, A., Roder, B., & Buchel, C. (2014a). Brain systems mediating voice identity processing in blind humans. *Human Brain Mapping, 35*(9), 4607–4619.

Holig, C., Focker, J., Best, A., Roder, B., & Buchel, C. (2014b). Crossmodal plasticity in the fusiform gyrus of late blind individuals during voice recognition. *Neuroimage, 103*, 374–382.

Holig, C., Focker, J., Best, A., Roder, B., & Buchel, C. (2017). Activation in the angular gyrus and in the pSTS is modulated by face primes during voice recognition. *Human Brain Mapping, 38*(5), 2553–2565.

Holmes, E., Domingo, Y., & Johnsrude, I. S. (2018). Familiar Voices Are More Intelligible, Even if They Are Not Recognized as Familiar. *Psychological Science, 29*(10), 1575–1583.

Holmes, E., & Johnsrude, I. S. (2021). Speech-evoked brain activity is more robust to competing speech when it is spoken by someone familiar. *Neuroimage, 237,* Article 118107. https://doi.org/10.1016/j.neuroimage.2021.118107

Hu, X., Wang, X., Gu, Y., Luo, P., Yin, S., Wang, L., . . . Chen, A. (2017). Phonological experience modulates voice discrimination: Evidence from functional brain networks analysis. *Brain and Language, 173*, 67–75.

Huang, Y., Huang, Q., Chen, X., Qu, T. S., Wu, X. H., & Li, L. (2008). Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults. *Hearing Research, 244* , 51–65.

Huber, E., Chang, K., Alvarez, I., Hundle, A., Bridge, H., & Fine, I. (2019). Early Blindness Shapes Cortical Representations of Auditory Frequency within Auditory Cortex. *Journal of Neuroscience, 39*(26), 5143–5152.

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. *Cerebral Cortex, 15*(10), 1609–1620.

Jonas, J., Jacques, C., Liu-Shuang, J., Brissart, H., Colnat-Coulbois, S., Maillard, L., & Rossion, B. (2016). A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proceedings of the National Academy of Sciences, 113*(28), E4088–E4097.

Jonas, J., Rossion, B., Brissart, H., Frismand, S., Jacques, C., Hossu, G., . . . Maillard, L. (2015). Beyond the core face-processing network: Intracerebral stimulation of a face-selective area in the right anterior fusiform gyrus elicits transient prosopagnosia. *Cortex, 72,* 140–155.

Kawashima, R., O'Sullivan, B. T., & Roland, P. E. (1995). Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the "mind's eye". *Proceedings of the National Academy of Sciences, 92*(13), 5969–5972.

Klinge, C., Eippert, F., Roder, B., & Buchel, C. (2010). Corticocortical connections mediate primary visual cortex responses to auditory stimulation in the blind. *Journal of Neuroscience, 30*(38), 12798–12805.

Koeda, M., Takahashi, H., Yahata, N., Asai, K., Okubo, Y., Tanaka, H. (2006). A functional MRI study: cerebral laterality for lexical-semantic processing and human voice perception. *Amercian Journal of Neuroradiology, 27*(7), 1472–1479.

Kupers, R., Fumal, A., de Noordhout, A. M., Gjedde, A., Schoenen, J., & Ptito, M. (2006). Transcranial magnetic stimulation of the visual cortex induces somatotopically organized qualia in blind subjects. *Proceedings of the National Academy of Sciences, 103*(35), 13256–13260.

Lane, C., Kanjlia, S., Richardson, H., Fulton, A., Omaki, A., & Bedny, M. (2017). Reduced Left Lateralization of Language in Congenitally Blind Individuals. *Journal of Neuroscience, 29*(1), 65–78.

Latinus, M., Crabbe, F., & Belin, P. (2011). Learning-induced changes in the cerebral processing of voice identity. *Cerebral Cortex, 21*(12), 2820–2828.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping, 24*(1), 11–20.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Neuroscience, 14*(3), 420–429.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge Eng. ; New York: Cambridge University Press.

Lewald, J., & Getzmann, S. (2013). Ventral and dorsal visual pathways support auditory motion processing in the blind: evidence from electrical neuroimaging. *European Journal of Neuroscience, 38*(8), 3201–3209.

Loiotile, R. E., Cusack, R., & Bedny, M. (2019). Naturalistic Audio-Movies and Narrative Synchronize "Visual" Cortices across Congenitally Blind But Not Sighted Individuals. *Journal of Neuroscience, 39*(45), 8940–8948.

Lomber, S. G., Meredith, M. A., & Kral, A. (2010). Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nature Neuroscience, 13*(11), 1421–1427.

Maguinness, C., Roswandowitz, C., & von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia, 116*, 179–193.

Maguinness, C., von Kriegstein, K. (2021). Visual mechanisms for voice-identity recognition flexibly adjust to auditory noise level. *Human Brain Mapping*, *42,* 3963–3982.

Manjunath, N. K., Srinivas, R., Nirmala, K. S., Nagendra, H. R., Kumar, A., & Telles, S. (1998). Shorter latencies of components of middle latency auditory evoked potentials in congenitally blind compared to normal sighted subjects. *International Journal of Neuroscience, 95*(3-4), 173–181.

Mathias, S. R., & von Kriegstein, K. (2014). How do we recognise who is speaking? *Frontiers in bioscience (Scholar edition), 6*, 92-109.

Matsuzaki, N., Juhasz, C., & Asano, E. (2012). Oscillatory modulations in human fusiform cortex during motion-induced blindness: intracranial recording. *Clinical Neurophysiology, 123*(10), 1925–1930.

Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., . . . Nestor, P. J. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain, 133*(11), 3256–3268.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., . . . Kojima, S. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia, 39*(10), 1047–1054.

Naveen, K. V., Srinivas, R. S., Nirmala, K. S., Nagendra, H. R., & Telles, S. (1997). Middle latency auditory evoked potentials in congenitally blind and normal sighted subjects. *International Journal of Neuroscience, 90*(1-2), 105–111.

Pang, W., Xing, H., Zhang, L., Shu, H., & Zhang, Y. (2020). Superiority of blind over sighted listeners in voice

recognition. *The Journal of the Acoustical Society of America, 148*(2), EL208–EL213.

Pant, R., Kanjlia, S., & Bedny, M. (2020). A sensitive period in the neural phenotype of language in blind individuals. *Developmental Cognitive Neuroscience, 41*, Article 100744. https://doi.org/10.1016/j.dcn.2019.100744

Parvizi, J., Jacques, C., Foster, B. L., Witthoft, N., Rangarajan, V., Weiner, K. S., & Grill-Spector, K. (2012). Electrical stimulation of human fusiform face-selective regions distorts face perception. *Jounral of Neuroscience, 32*(43), 14915–14920.

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research, 134*, 427–445.

Pelland, M., Orban, P., Dansereau, C., Lepore, F., Bellec, P., & Collignon, O. (2017). State-dependent modulation of functional connectivity in early blind individuals. *Neuroimage, 147*, 532–541.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117*(6), 1283–1301.

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., . . . Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage, 119*, 164–174.

Poirier, C., Collignon, O., Scheiber, C., Renier, L., Vanlierde, A., Tranduy, D., . . . De Volder, A. G. (2006). Auditory motion perception activates visual motion areas in early blind subjects. *Neuroimage, 31*(1), 279–285.

Ptito, M., Fumal, A., de Noordhout, A. M., Schoenen, J., Gjedde, A., & Kupers, R. (2008). TMS of the occipital cortex induces tactile sensations in the fingers of blind Braille readers. *Experimental Brain Research, 184*(2), 193–200.

Qin, W., & Yu, C. (2013). Neural pathways conveying novisual information to the visual cortex. *Neural Plasticity, 2013,* Article 864920. https://doi.org/10.1155/2013/864920

Rama, P., & Courtney, S. M. (2005). Functional topography of working memory for face or voice identity. *Neuroimage, 24*(1), 224–234.

Ratan Murty, N. A., Teng, S., Beeler, D., Mynick, A., Oliva, A., & Kanwisher, N. (2020). Visual experience is not necessary for the development of face-selectivity in the lateral fusiform gyrus. *Proceedings of the National Academy of Sciences, 117*(37), 23011–23020.

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Current Biology, 21*(5), 363–368.

Rimmele, J. M., Gudi-Mindermann, H., Nolte, G., Roder, B., & Engel, A. K. (2019). Working memory training integrates visual cortex into beta-band networks in congenitally blind individuals. *Neuroimage, 194*, 259–271.

Roder, B., Kramer, U. M., & Lange, K. (2007). Congenitally blind humans use different stimulus selection strategies in hearing: an ERP study of spatial and temporal attention. *Restorative Neurology and Neuroscience, 25*(3-4), 311–322.

Roder, B., Rosler, F., Hennighausen, E., & Nacker, F. (1996). Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects. *Cognitive Brain Research, 4*(2), 77–93.

Roder, B., Rosler, F., & Neville, H. J. (1999). Effects of interstimulus interval on auditory event-related potentials in congenitally blind and normally sighted humans. *Neuroscience Letters, 264*(1-3), 53-56.

Schall, S., Kiebel, S. J., Maess, B., & von Kriegstein, K. (2013). Early auditory sensory processing of voices is facilitated by visual mechanisms. *Neuroimage, 77*, 237–245.

Schall, S., Kiebel, S. J., Maess, B., & von Kriegstein, K. (2015). Voice identity recognition: functional division of the right STS and its behavioral relevance. *Journal of Neuroscience, 27*(2), 280–291.

Sharda, M., & Singh, N. C. (2012). Auditory perception of natural sound categories--an fMRI study. *Neuroscience, 214*, 49–58.

Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience, 11*(11), 1812–1822.

Schepers, I. M., Hipp, J. F., Schneider, T. R., Roder, B., & Engel, A. K. (2012). Functionally specific oscillatory activity correlates between visual and auditory cortex in the blind. *Brain, 135*(3), 922–934.

Schweinberger, S. R. (2001). Human brain potential correlates of voice priming and voice recognition. *Neuropsychologia, 39*(9), 921–936.

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zaske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(1), 15–25.

Schweinberger, S. R., Walther, C., Zaske, R., & Kovacs, G. (2011). Neural correlates of adaptation to voice identity. *British Journal of Psychology, 102*(4), 748–764.

Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science, 366*(6461), 58-62.

Shulman, G. L., Corbetta, M., Buckner, R. L., Raichle, M. E., Fiez, J. A., Miezin, F. M., & Petersen, S. E. (1997). Top-down modulation of early sensory cortex. *Cerebral Cortex, 7*(3), 193–206.

Stevenage, S. V. (2018). Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. *Neuropsychologia, 116*, 162–178.

Topalidis, P., Zinchenko, A., Gadeke, J. C., & Focker, J. (2020). The role of spatial selective attention in the processing of affective prosodies in congenitally blind adults: An ERP study. *Brain Research, 1739*, Article 146819. https://doi.org/10.1016/j.brainres.2020.146819

Van Lancker, D., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A Dissociation Between Familiar and Unfamiliar Voices. *Cortex, 24*(2), 195–209.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25*(5), 829–834.

Van Lancker, D., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology, 11*(5), 665–674.

von Kriegstein, K., Dogan, O., Gruter, M., Giraud, A. L., Kell, C. A., Gruter, T., . . . Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences, 105*(18), 6747–6752.

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research, 17*(1), 48–55.

von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus

for the processing of voices. *Neuroimage, 22*(2), 948–955.

von Kriegstein, K., Kleinschmidt, A., & Giraud, A. L. (2006). Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cerebral Cortex, 16*(9), 1314–1322.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Neuroscience, 17*(3), 367–376.

Voss, P., Gougoux, F., Zatorre, R. J., Lassonde, M., & Lepore, F. (2008). Differential occipital responses in early- and late-blind individuals during a sound-source discrimination task. *Neuroimage, 40*(2), 746–758.

Voss, P., Lepore, F., Gougoux, F., & Zatorre, R. J. (2011). Relevance of spectral cues for auditory spatial processing in the occipital cortex of the blind. *Frontiers in Psychology, 2*, Article 48. https://doi.org/10.3389/fpsyg.2011.00048

Warren, J. D., Scott, S. K., Price, C. J., & Griffiths, T. D. (2006). Human brain mechanisms for the early analysis of voices. *Neuroimage, 31*(3), 1389–1397.

Watkins, K. E., Shakespeare, T. J., O'Donoghue, M. C., Alexander, I., Ragge, N., Cowey, A., & Bridge, H. (2013). Early Auditory Processing in Area V5/MT+ of the Congenitally Blind Brain. *Journal of Neuroscience, 33*(46), 18242–18246.

Wilms, M., Eickhoff, S. B., Specht, K., Amunts, K., Shah, N. J., Malikovic, A., & Fink, G. R. (2005). Human V5/MT+: comparison of functional and cytoarchitectonic data. *Anatomy and Embryology, 210*(5-6), 485–495.

Xu, L. J., Li, J. Y., Wu, X. H., & Li, L. (2008). Modulation of the voice-cuing effect on releasing speech from informational masking. *Journal of the Acoustical Society of America, 123,* 3864.

Yang, Z. G., Chen, J., Huang, Q., Wu, X. H., Wu, Y. H., Schneider, B. A., & Li, L. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication, 49,* 292–904.

Zaske, R., Awwad Shiekh Hasan, B., & Belin, P. (2017). It doesn't matter what you say: FMRI correlates of voice learning and recognition independent of speech content. *Cortex, 94*, 100–112.

Zhang, Z. J., Hao, G. F., Shi, J. B., Mou, X. D., Yao, Z. J., & Chen, N. (2008). Investigation of the neural substrates of voice recognition in Chinese schizophrenic patients with auditory verbal hallucinations: an event-related functional MRI study. *Acta Psychiatrica Scandinavica, 118*(4), 272–280.

The neural mechanisms for human voice processing
—— Neural evidence from sighted and blind subjects

MING Lili[1,2], HU Xueping[1,2,3]

([1] School of Linguistic Science and Art, Jiangsu Normal University, Xuzhou 221000, China) ([2] Key Laboratory of Language and Cognitive Neuroscience of Jiangsu Province, Collaborative Innovation Center for Language Ability, Xuzhou 221009, China) ([3] School of Chinese Language and Culture, Nanjing Normal University, Nanjing 210097, China)

Abstract: The human voice, as an important part of one's auditory environment, contains a large

amount of paralinguistic information to help identify individuals. Especially for blind individuals, the lack of visual face experience makes voice information the main source of perceiving another person's individual characteristics. By combining the research on voice-selective and voice-identity processing among both sighted and blind individuals, the present study attempts to reveal the general human voice processing mode and specific voice processing mechanism among blind persons. It also shows that the fusiform gyrus related to visual face processing are involved in voice tasks in both groups. The auditory-visual integration model and cross-modal reorganization hypothesis based on the "unmasking effect", provide respective explanation mechanisms for the different neural bases of voice processing between the two subjects. Future, researchers should further investigate the voice processing strategies of blind individuals and the role of the left superior temporal sulcus/gyrus in voice processing.

Keywords: voice identity processing, sighted subjects, blind subjects, fusiform gyrus, face processing